DOCUMENT RESUME

ED 229 390 TM 830 202

AUTHOR Marklund, Sixten
TITLE Applicability of "Standards for Evaluations of Educational Programs, Projects and Materials" in an International Setting.
PUB DATE Apr 83
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Change; Elementary Secondary Education; Evaluation Methods; Foreign Countries; *International Studies; *Program Evaluation; Reading Comprehension; Science Education; *Standards
IDENTIFIERS *Standards for Evaluation Educ Prog Proj Materials; Sweden (Stockholm)

ABSTRACT

This paper is a reaction to the "Standards for Evaluations of Educational Programs, Projects and Materials," focusing on the applicability of these standards. The evaluation of programs almost without exception deals with the evaluation of change of programs, taking the form of comparative studies, and being valued according to how the program resembles or deviates from other programs. The actual foundations for a program evaluation are different on both sides of the Atlantic. A set of standards in America is only partly applicable in Europe. Despite the clear, well thought-out, and systematic structure of these standards, the value of them outside of the United States stands the risk of being limited to becoming mainly of technical nature. Evaluation examples of educational programs (reading comprehension and science education) show how data from a program evaluation can be used in different ways and be interpreted differently according to the different educational political goals. The conclusion of a revision of standards is that these are valuable and should always be aimed at, but that they cannot conceal the diversions in educational political values, which are the great problems in the evaluations. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED229390

1983-03-15

APPLICABILITY OF "STANDARDS FOR EVALUATIONS OF EDUCATIONAL PROGRAMS, PROJECTS AND MATERIALS" IN AN INTERNATIONAL SETTING

By Sixten Marklund
Institute of International Education
University of Stockholm, Sweden

Contents

	Page
Summary	1
1. Standards for Evaluation of Educational Programs	2
2. Evaluation of Programs and Evaluation of Changes	
3. Applicability of Standards	5
4. Reading Comprehension	7
4.1 The "Olympic Games" in Reading Comprehension	7
4.2 Alternative Evaluation of Reading Ability	8
5. Science	12
5.1 Mean Result as Criterion	12
5.2 Top Results as Criterion	13
6. Good and Bad Schools	14
7. Back to "Standards for the Evaluation of Educational Programs"	17
References	19

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

~~W.B. Dockrell~~
S. MARKLUND

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830202

1. Standards for Evaluation of Educational Programs

This preliminary paper is a spontaneous reaction to the book "Standards for Evaluations of Educational Programs, Projects and Materials". Reading the book has been very rewarding to me. It is the result of extensive work of experts. The standards presented have been carefully discussed and have been given exemplary structure and presentation. The system of standards is simple and clear. It is uniform from one group of standards to another. The book is well-arranged and readable.

The authors of these standards are undoubtedly aware of the risks of giving their standards such a well-arranged, uniform and readable structure. I have taken it as my task here to call attention to some points about the applicability of these standards. I do this based on the experience that school systems across the world are different and that they differ not only in structure but also in the goals they are a manifestation of. Moreover, one school system (it may also apply to what in the book is called "program") can be valued differently by those it is intended to serve. What one person sees as merits of the program can in the eyes of the other be the deficiency of the program.

By contrast one can say, like the author of the book, that if the goals and objectives of a program really have been given a clear formulation, there should be no disagreement on this, and consequently there should be no disagreement on how it should be evaluated and how the evaluation results should be interpreted either. Unfortunately it is often not that simple. It is rather the exception than the rule that the goals are clear. Agreement is reached on uniformly formulated goals, despite differing values, often only at the cost of vague and amgibuous goal descriptions. The ambiguity that can be

concealed through generally formulated compromise decisions on the goals, almost inevitably appear when in the end the results of the evaluation are to be interpreted.

An initial question that I ask myself when reading the book is if what is called "programs, projects, and materials" really can be evaluated in the same way, i.e. if the same standards can be used for these three areas. As I see it, projects and materials can be regarded as a separate group of evaluation objects. They are usually clearly defined with respect to time and content. They have also been created for reasonably clearly specified objectives. By contrast, one can put educational programs where goals and objectives have a significantly larger range. Programs have to be regarded as including not only specially destined courses, but also the general local, regional, and national educational systems that are a result of an encompassing educational policy. They apply for a long time. They also apply to large groups of individuals in society. They form the base for the specialized programs (perhaps called projects and materials), that by experience are easier to evaluate according to the standards given in the book.

2. Evaluation of Programs and Evaluation of Changes

I restrict my points of view here to educational programs in the general and encompassing sense I have suggested, and also to the experiences in such educational programs in Western Europe, perhaps primarilly in Scandinavia. Great changes in the educational programs have been made here during the last decades. The evaluation could be limited to restricted programs, their function, and results. Usually it is not the function and results of these restricted programs that are asked for at the evaluation, but rather the results from the educational reforms at large.

A first such reform is the prolongation of compulsory schooling. Where there normally was six or seven years' compulsory schooling earlier, it has in a comparatively short period of time been prolonged to eight, nine, or sometimes ten years' compulsory schooling. How should this prolongation be evaluated?

A second reform recently that perhaps has drawn more attention educational policy-wise, is that the opportunities for post-compulsory schooling have increased markedly. Whereas only 20 or 30 per cent of each age cohort could continue to post-compulsory schooling a couple of decades ago, now more than half continue, in some countries as much as 80 or 90 per cent. This includes not only vocational training. The paths to so called higher academic education have also been considerably broadened. How shall standards for evaluation be formulated to help evaluations of these changes and of the new educational systems they have led to?

As regards evaluation of educational programs, a third change made in post-war Western Europe is equally important, namely that previously separate educational programs (school systems), have been coordinated within a uniform frame to so called comprehesive school programs. Programs that earlier have been socio-economically clearly segregated, have been merged through political decisions. Lower secondary school has become a part of the comprehensive school. How should such a program change be evaluated? Could it be carried out according to the standards given in the book?

Closely related to this program change is a fourth and in terms of evaluation equally important, namely that the way in which the choice of line of studies and of specialization after comprehensive compulsory schooling occurs, has changed during recent years. Whereas earlier it was the school which distributed students over different educational paths, primarily through a division on academic versus non-academic, it is now usually done at the sudent's own choice. It is the student and his parents who choose continued education, not the school and teachers who differentiate the students. This is a change that strikes differently in different school systems (school programs) and it is important to evaluate its forms and effects.

To this can be added a change of a fifth kind. The distribution of the students over different study paths now occurs later than it did earlier. Whereas the choice to so called higher studies, i.e. to academic courses, previously was done at the age of 10-12, it is now postponed until the age of 14-16. This postponement has in many ways proved to have great significance not only for the students themselves but also for programs of higher education as well as programs of

vocational education, for industry and for society at large.

Thus we find that the evaluation of programs almost without exception deals with the evaluation of change of programs. It concerns the questions that are being posed after these changes by not only the individual but also by those responsible for societal development at large: Did it get better or worse as a result of the changes? If so, for whom and in what respects? It is in the nature of things that program evaluations of this kind almost always take the form of comparative studies. A program is valued according to how it resembles or deviates from other programs, earlier or contemporary.

The evaluation issues can be of technical and limited kind. This makes standards and evaluation criteria easier to determine. Experience says, however, that the important and difficult issues almost always are in some respect political. They deal with social, economic, and cultural problems. Early or late differentiation, comprehensive school system or parallel school system, free choice of specialization or selection of students by school, schooling with or without fees, economic support to all, to those in special need or to no-one, etc., have lately been political issues in Europe. Conservative, liberal, and socialist parties here differ in their values. The differences are rather permanent and rather similar from country to country. In American educational policy this left-right dimension is less permanent and less clear. The differences here often manifest themselves within the political parties, not between them. They vary there from case to case.

3. Applicability of Standards

The difference between Europe and America regarding the view on educational politics and educational programs can be explained historically. We shall here abstain from giving such explanations. The essential thing is that the actual foundations for a program evaluation are different on both sides of the Atlantic. Thereby a set of standards in America is only partly applicable in Europe. Despite the clear, well thought-out, and systematic structure of these standards, the value of them outside of the U.S. stands the risk of being limited to becoming mainly of technicla nature. Already in the first of the 30 standards given in the publication, i.e. "Audience

English and French as foreign languages. The applicability of different "standards for the evaluation of educational programs", in this case the national school systems in the participating countries, can be illustrated with some of the knowlege evaluations. (We here abstain from describing the complicated procedure of constructing the international knowlege tests, the sampling of schools and students, the relations of the results to non-cognitive outcomes, school organization and teaching methods).

4. Reading Comprehension

4.1 The "Olympic Games" in Reading Comprehension

The examples of evaluation of educational programs given below are taken from the IEA-project. They do not aim at an evaluation of the school systems in the countries that were included in this project, but only at showing how data from a program evaluation can be used in different ways and be interpreted differently according to the different educational political goals.

A first example concerns reading comprehension in 15 countries for 10-, and 14-year old students. Means (M) and standard deviations (SD) for a reading test of 45 and 52 items respectively, turned out as shown in Table 1.

The purpose of the project was naturally not restricted to comparing the means of the countries. The relations in the results to the structure of the educational system and to the students' background were more interesting from a research point of view. Nevertheless the interest aroused by the results has to a great extent been connected just with a comparison of the means. It became the "Olympic Games of Knowlege". In medal positions for 10-year olds came Sweden, England, and Finland. For the 14-year olds it was New Zealand, United States, and Italy. Chile, India, and Iran formed a separate group, for sake of simplicity referred to here as developing countries, with clearly lower results than the others. The latter, referred to here as industrialized countries, did not greatly differ internally in mean results.

Table 1. Reading Comprehension in Fifteen Countries for 10 and 14 Year Old Students. Individual Results.[1)]

Country	10 Year Olds		14 Year Olds	
	M	SD	M	SD
Belgium / Fl[2)]	24.1	6.9	31.0	7.0
Belgium / Fr[2)]	24.4	7.2	33.6	6.6
Chile	17.6	7.4	23.1	8.2
England	24.9	8.8	31.7	9.0
Finland	25.4	8.1	33.2	8.3
Hungary	21.5	7.4	32.1	7.4
India[3)]	17.8	7.3	16.4	5.7
Iran	13.0	5.7	18.1	5.2
Israel	22.1	8.6	31.0	9.3
Italy	-	-	33.7	6.9
Netherlands	24.1	7.2	31.6	7.8
New Zealand	-	-	34.7	8.3
Scotland	24.3	9.0	32.9	8.6
Sweden	27.1	8.1	32.1	8.1
United States	23.8	8.7	34.3	8.4

1) The statistical data given here and in the following tables are calculated on the basis of the IEA data bank at the Institute of International Education, University of Stockholm.

2) Belgium is treated as two "nations", a Flemish and a French.

3) Only Hindi speaking students

4.2 Alternative Evaluation of Reading Ability

As shown in Table 1, the reading results of the countries also differ in terms of distribution (SD). In a typical ability subject such as reading, the distribution downwards from the mean is of special interest. In the same way as an aeroplane needs a certain speed to lift, the student needs a certain reading ability to profit from the reading. A crucial issue here is where the minimum competence lies, i.e. the result in the test that everybody should achieve. Functional reading ability is one of the goals of the program, perhaps the most

Table 2. Reading Comprehension for 10 and 14 Year Old Students in Fifteen Countries. Individual Results. Proportion of Students below "Minimum Scores" (Points p).

Country	% of Students below "Minimum Scores"			
	10 Year Olds		14 Year Olds	
	11 p	16 p	16 p	21 p
Belgium /Fl	1.4	11.9	3.2	7.1
Belgium /Fr	3.5	12.4	0.7	3.1
Chile	17.5	46.0	19.9	42.1
England	5.3	17.7	4.5	12.3
Finland	4.0	14.6	2.3	8.8
Hungary	6.2	24.7	1.5	6.8
India	17.4	42.4	48.6	78.8
Iran	37.3	75.4	31.2	69.8
Israel	9.7	27.5	6.0	16.3
Italy	-	-	0.7	3.6
Netherlands	3.2	13.5	1.4	8.5
New Zealand	-	-	2.2	6.8
Scotland	5.2	19.0	3.0	9.0
Sweden	2.4	11.1	2.9	10.1
United States	5.8	21.5	3.0	7.3

important. That the individual gradually improves and reaches above this minimum level is of course desirable but not decisive in the same way when the program should be evaluated. The over-achievement of the one can never compensate for the achievement under the minimum level of the other.

In Table 2 is indicated how many per cent of the 10-year olds in the different countries fall under 11 or 16 points respectively in the test. Just these breaking points can of course be discussed, which we here abstain from. Our primary interest is only to see how a program evaluation of this kind differs from the previous one, i.e. the comparison of the means of Table 1. Apparently it becomes here another kind of "Olympic Games". Some countries with comparatively high means, but at the same time with wide distribution, now come lower.

Table 3. Reading Comprehension. Countries Ranked According to Three Criteria

	10 Year Old Students		
	M_{11}	< 11 points	< 16 points
1	Sweden	Belgium / Fl	Sweden
2	Belgium / Fr	Sweden	Belgium / Fl
3	Finland	Netherlands	Belgium / Fr
4	England	Belgium / Fr	Netherlands
5	Scotland	Finland	Finland
6	Belgium / Fl	Scotland	England
7	Netherlands	England	Scotland
8	United States	United States	United states
9	Israel	Hungary	Hungary
10	Hungary	Israel	Israel
11	India	India	India
12	Chile	Chile	Chile
13	Iran	Iran	Iran

	14 Year Old Students		
	M_{14}	< 16 points	< 21 points
1	New Zealand	Belgium / Fr	Belgium / Fr
2	United States	Italy	Italy
3	Sweden	Netherlands	New Zealand
4	Italy	Hungary	Hungary
5	Belgium / Fr	New Zealand	Belgium / Fl
6	Scotland	Finland	United States
7	Netherlands	Sweden	Netherlands
8	Finland	United States	Finland
9	Hungary	Scotland	Scotland
10	England	Belgium / Fl	Sweden
11	Israel	England	England
12	Belgium / Fl	Israel	Israel
13	Chile	Chile	Chile
14	Iran	Iran	Iran
15	India	India	India

Reading ability is a "must" in modern society. For his own well-being and for his roll as an active member of society, the individual must have at least a minimum level of competence. The results in Table 2 are therefore more essential than the results in Table 1. There is nothing to prevent making a program evaluation in both ways, using this kind of data.

If we return to the idea of evaluation as an "Olympic Games" in reading ability, we find that the countries rank differently in regard to the mean and to the minimum demand. In Table 3 this has been done for the 10-year and 14-year olds. The left column corresponds to the ranking by means in Table 1, the two other rank orders are according to the two minimum criteria in Table 2.

Table 4. Science Test Scores in Eighteen Countries for 10 and 14 Year Old Students. Individual Results.

Country	10 Year Olds		14 Year Olds	
	M	SD	M	SD
Australia	-	-	34.2	10.0
Belgium / Fl	21.7	5.9	31.7	7.0
Belgium / Fr	18.9	5.8	25.8	8.2
Chile	14.5	7.2	20.6	7.9
England	20.3	7.0	32.0	11.6
Finland	21.6	6.6	31.4	8.7
Germany	19.6	6.2	33.7	9.5
Hungary	21.2	6.7	39.0	10.3
India	14.2	6.6	20.6	7.9
Iran	10.4	4.7	19.9	5.5
Italy	-	-	28.1	8.3
Japan	25.2	6.2	40.6	11.8
Netherlands	19.7	6.1	27.4	10.0
New Zealand	-	-	34.6	10.7
Scotland	18.6	7.5	32.1	11.3
Sweden	22.2	6.0	32.7	9.5
Thailand	15.5	5.3	28.4	6.4
United States	22.2	7.4	33.7	9.4

Apparently most countries change their rank order when the criterion changes, which can be seen from the lines connected with the individual country in the three rank orders. Thus, Sweden ranks 1-2-1, Belgium/Fr. ranks 2-4-3, Finland 3-5-5, etc.. Only in the lower parts of the lists, dominated by the developing countries, is the rank order more constant.

A corresponding comparison for the 14-year olds gives a similar reslult, which can be seen in the lower part of the Table.

5. Science

5.1 Mean Result as Criterion

In the IEA-project science was evaluated in the same way as reading comprehension, i.e. on the basis of mean (M) and standard deviation

Table 5. Science Achievement. Proportion of 10 Year Old Students Passing 25 Points and 14 Year Old Students Passing 45 Points. Individual Results.

10 Year Olds		14 Year Olds	
Country	% Students Passing 25 p	Country	% Students Passing 45 p
Japan	52.1	Japan	34.2
USA	36.6	Hungary	26.2
Finland	30.4	Australia	26.2
Sweden	30.3	New Zealand	15.9
Hungary	28.4	Scotland	13.3
Belgium / Fl	28.2	England	13.2
England	24.8	USA	11.7
Netherlands	18.9	Germany	10.7
Germany	16.5	Sweden	10.2
Scotland	16.5	Finland	6.9
Belgium / Fr	13.9	Belgium / Fl	3.4
Chile	8.2	Netherlands	3.3
India	7.0	Italy	3.2
Thailand	3.7	Chile	1.5
Iran	0.5	India	1.4
		Belgium / Fr	1.3
		Thailand	0.9
		Iran	0.0

(SD) for each country. The test encompassed 40 items for the 10-year olds and 80 items for the 14-year olds. The results can be seen in Table 4.

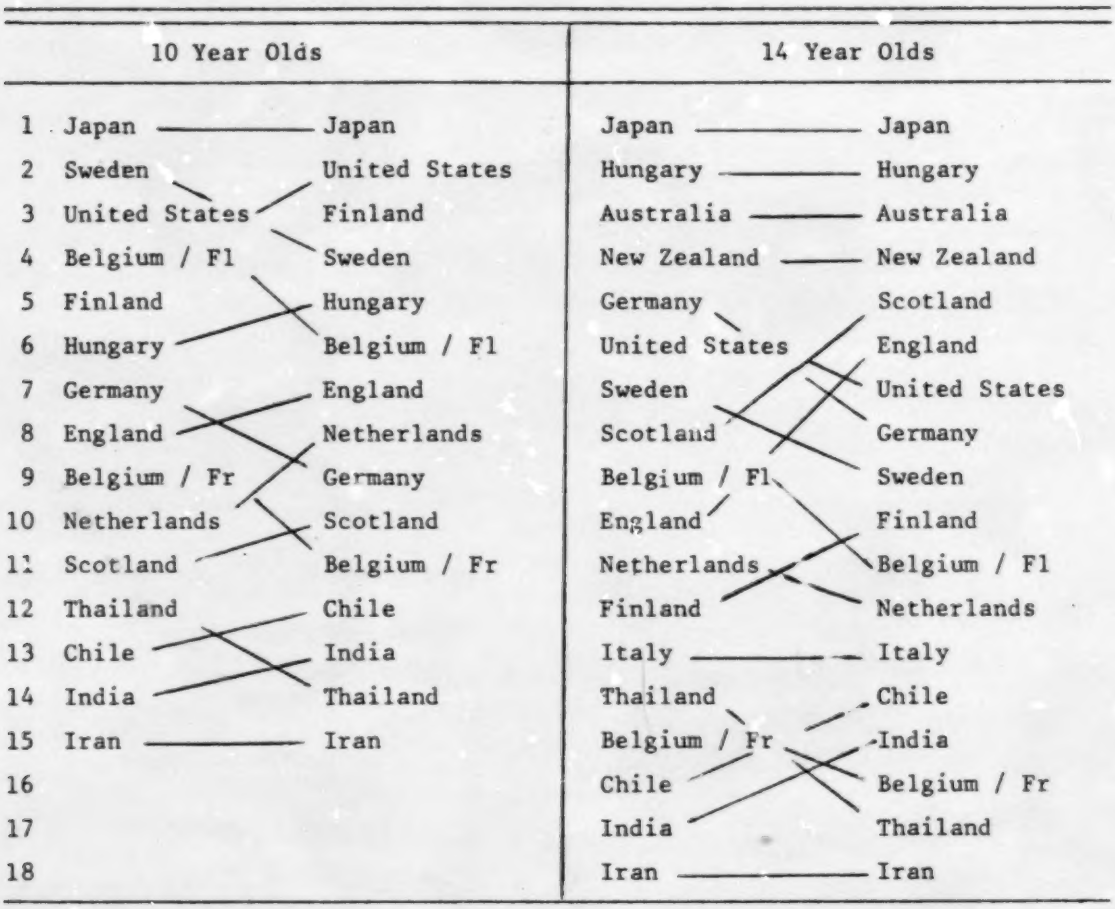
As can be seen in the Table, the best results in both age groups are achieved by Japan, which did not participate in reading comprehension. Chile, India, and Iran also here form a separate group with lower results than the others, which thus lie higher than the developing countries but lower than Japan. In"medal positions" for the 10-year olds come Japan, Sweden, and the United States. For the 14-year olds there are Japan, Hungary, and Australia. Many differences between the countries are small.

5.2 Top Results as Criterion

As for reading comprehension a minimum level can here be defined, and the extent to which different countries reach this level can be evaluated. In science, however, such a lower limit is not as important as in reading comprehension. It is also more difficult in science to reach an agreement on where it should be drawn. It can, however, be interesting, for a general subject like science, to see which countries reach top results. Not only the individual himself benefits from good knowlege. Society and industry at large benefit from there being competent people in this area. It can then be interesting to compare the proportion of the population above a certain limit. Such a comparison has been made in Table 5. The limit has there been put at 25 points for the 10-year olds and at 45 pionts for the 14-year olds. The countries have been rank ordered according to the proportion of high achieving students.

Also here Japan tops the lists. For the 14-year olds follow then Hungary, Australia, and New Zealand, i.e. the same countries as in the rank order according to mean (Table 4). Most other countries, though, get new positions, which can be seen in Table 6. The pattern of lines connecting individual countries, tend to resemble the corresponding pattern for the comparison between mean and minimum result, although in this case - which is not surprising - becomes turned upside-down (compare the 14-year olds in Table 6 with the 14-year olds in Tabel 3).

Table 6. Science Achievement. Countries Ranked According to Mean Score (left) and Top Results (right). Individual Results.



6. Good and Bad Schools

The previous evaluations have dealt with the results at the student level. With the same data a corresponding evaluation can be carried out at the school level or at the school district level. The countries can achieve equally in terms of means or frequences of low or high results, but still differ in how these low and high results are distributed over individual schools. The goal of equal opportunity implies that the schools should have an even standard. Students' and parents' hunting for admission to "good schools" is well-known.

For educational politicians and administrators of education, the problem here is tied to the presence of "weak" schools. Experience

tells us that they get even weaker if nothing is done to them. Good teachers try to get away from them. The parents look for better schools for their children. As a result the grants to the schools decrease, which further worsens their results in a downward spiral of causes and effects.

Table 7. Reading Comprehension for 10 Year Old Students. Frequency of Schools with Means under 16 and 21 Points

% of Schools with M < 16		% of Schools with Mean < 21	
Belgium / Fl	0.0	Sweden	6.0
Sweden	0.0	Finland	6.6
Finland	1.2	Netherlands	9.0
Netherlands	1.5	England	11.3
Scotland	1.8	Scotland	14.4
Belgium / Fr	3.0	Belgium / Fl	21.1
England	3.1	Belgium / Fr	21.2
United States	5.2	United States	22.6
Hungary	5.9	Hungary	48.6
Israel	17.9	India	55.6
Chile	42.7	Israel	55.6
India	51.1	Chile	70.9
Iran	70.7	Iran	87.7

The mean result in the individual school is therefore essential in a program evaluation. Table 7, which refers to reading comprehension for 10-year olds, gives the frequences of the schools in the different countries with mean results under 16 and 21 points respectively. Also these points taken as "the lowest acceptable school level" can of course be discussed. What the Table shows is precisely to what extent low student results are concentrated to certain schools. The higher the percentages are, the more uneven is the school standard.

With 16 as the lower "school limit", Belgium/Fr. and Sweden do apparently not have any "bad" schools. For the other industrialized countries their proportion increases to 17.9 per cent (Israel). For the developing countries the proportion is clearly higher.

Table 8. Science Achievement for 14 Year Olds. Frequency of Schools with Means under 29 Points

Country	% of Schools
Japan	0.0
Hungary	2.1
Australia	7.2
New Zealand	8.3
Sweden	9.3
Germany	12.6
United States	20.3
Finland	36.4
Scotland	37.0
Belgium / Fl	41.4
Netherlands	44.6
England	45.7
Belgium / Fr	61.9
Thailand	62.8
Italy	63.0
India	87.9
Chile	90.9
Iran	100.0

Hungary, Australia, New Zealand, Sweden, and Germany don't come too far behind Japan. There are considerably more "weak" schools in the United States, Finland, Scotland, Belgium/Fl., Netherlands, and England, the latter two with almost half of the schools under the level of Japan's lowest school. Of the last countries on the list, the majority have a school mean under the lowest one of Japan. It may be noted that the best school of Iran lies lower than the worst school in Japan.

The rank order is apparently not considerably changed, if the limit is raised to 21 points. The two rank orders should be compared with the rank order on the student level according to Table 1. They differ. The fact that the results of the students in the individual country vary from high to low, can in varying degrees be attributed to the fact that the schools within the individual country differ from each other.

In school politics of most countries the ambition has been to decrease the between school variance through postponed differentiation and through different kinds of support. The less the between school variance is as a proportion of the total variance of the country, the more "equal opportunity" is achieved.

Also in science the schools differ in terms of the size of the between school variance, which is confirmed in Table 8. There 29 points has been set as an acceptable school mean for 14-year olds. This is the lowest school mean for Japan, and as we have seen from Tables 4-6, the highest achieving contry in science, regardless of the kind of criterium. Naturally also the limit at 29 points can be discussed. But exactly as it in the argumentation for a minimum limit on the individual level in reading comprehension was suggested important with such a limit, it is also here important with such a limit. Schools in the public educational system must simply not get "bad". They have to satisfy a minimum demand to be acceptable and to avoid the risk of gradually declining further.

7. Back to "Standards for the Evaluation of Educational Programs"

With these examples of program evaluation I have wanted to show that the application of standards helps the evaluating of various programs by defining goals, means, and results. Standards is a useful checklist of the demands that can be placed on an evaluation, and they ought to be followed. But one should also be aware that through the application of them the differences in value pattern which lie behind the evaluation are not eliminated. It is said in the book: "The standards do not encompass a particular view of what constitutes good education, nor do they present specific criteria by which to judge

educational programs, projects, and materials." The joint committee also says it wants to "encourage the sound use of a variety of evaluation methods" (p. 7).

An evaluation of programs - as it can always be said - hereby gets strongly dependent on what one as an individual and member of society wants to accomplish with this program. It becomes a political issue, at least an educational political issue.

Here the problems of equality and varying definitions of it arise. Here also the problems of the need for experts and over-achieving students arise. There must be those who achieve more than others. Also the problem with how the weak and under-achieving should be fit into our community becomes apparent. The conclusion of a revision of standards is that these are valuable and should always be aimed at, but that they cannot conceal the diversions in educational political values, which are the great problems in the evaluations. Evaluations should be kept separate on two levels: one technical where standards always are desireable and important, and one political where they are disirable, but not decisive. Here the values take over. What we <u>want</u> with school then means more than what we <u>know</u> about it. At the same time it is true - as always - that the work with "how school should be" is facilitated when we know "how school is".

References

Comber, L C & Keeves, J P (1973). Science Education in Nineteen Countries. International Studies in Evaluation I. International Association for the Evaluation of Educational Achievement. Almqvist & Wiksell, Stockholm.

The IEA Data Base. The International Association for the Evaluation of Educational Achievement. Institute of International Education, University of Stockholm, Sweden.

The Joint Committee on Standards for Educational Evaluation (1981). Standards for Evaluations of Educational Programs, Projects, and Materials. McGraw-Hill Book Company, New York.

Thorndike, R L (1973). Reading Comprehension Education in Fifteen Countries. International Studies in Evaluation III. International Association for the Evaluation of Educational Achievement. Almqvist & Wiksell, Stockholm,